

University of Groningen

## **Why do faultlines matter? A computational model of how strong demographic faultlines undermine team cohesion**

Flache, Andreas; Mas, Michael; Mäs, Michael

*Published in:*  
Simulation Modelling Practice and Theory

*DOI:*  
[10.1016/j.simpat.2007.11.020](https://doi.org/10.1016/j.simpat.2007.11.020)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2008

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

Flache, A., Mas, M., & Mäs, M. (2008). Why do faultlines matter? A computational model of how strong demographic faultlines undermine team cohesion. *Simulation Modelling Practice and Theory*, 16(2), 175-191. <https://doi.org/10.1016/j.simpat.2007.11.020>

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# Why do faultlines matter? A computational model of how strong demographic faultlines undermine team cohesion

Andreas Flache\*, Michael Mäs

*Department of Sociology/ICS, University of Groningen, The Netherlands*

Available online 1 February 2008

---

## Abstract

Lau and Murnighan (LM) suggested that strong demographic faultlines threaten team cohesion and reduce consensus. However, it remains unclear which assumptions are exactly needed to derive faultline effects. We propose a formal computational model of the effects of faultlines that uses four elementary social mechanisms, social influence, rejection, homophily and heterophobia. We show that our model is consistent with the central hypotheses of LM's theory. We also find that negative effects of faultlines can be derived even when – unlike LM – we assume that initially there is no correlation between the demographic characteristics and the opinions of team members.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** Demographic diversity; Demographic faultline; Social cohesion; Work team; Agent-based modeling

---

## 1. Introduction

Immigration and the internationalization of organizations are trends that make work teams in modern work organizations increasingly diverse in terms of their demographic composition. This led management scholars to study intensively the consequences of diversity on team performance. However, diversity research has not produced a clear cut picture. While some authors reported positive effects, others found that more demographic diversity may reduce team performance (for comprehensive reviews about theoretical and empirical research see: [3,29,31,44,47,49]). In search for explanatory mechanisms, researchers have identified *two main arguments* [38]. Diversity can be beneficial for organisations because diverse teams have a larger pool of *social and human capital* that can be used to increase the teams' performance (e.g. [7]). But at the same time, diverse teams tend to be less *cohesive* than homogenous teams and cohesion is a main determinant of performance (e.g. [18,19]). The less cohesive a team is the less will it be able to make use of its social and human capital. Milliken and Martins concluded that “diversity thus appears to be a double-edged sword” [29, 403].

Milliken and Martins' conclusion inspired researchers to search for conditions that mediate the link between diversity and cohesion. Lau and Murnighan [23,24] in particular added a new idea to the debate. They

---

\* Corresponding author.

E-mail address: [A.Flache@rug.nl](mailto:A.Flache@rug.nl) (A. Flache).

suggested that it may not be the level of demographic diversity that puts cohesion under pressure, but it may be the way diverse demographic attributes are distributed in a team. More in particular, in their view cohesion suffers in a diverse group only to the extent that the distribution of attributes across group members generates a *strong demographic faultline*. A demographic faultline exists if not all members of the team are similar with respect to a salient demographic attribute, but only strong faultlines cause problems. Here and in the remainder of this paper, we consider those individual attributes as ‘demographic’ which are either fixed or change only so slowly (such as educational level) that they can be considered as constant in the time frame in which opinion formation in groups takes place. Group faultlines increase in strength as more attributes are highly correlated, reducing the number and increasing the homogeneity of resulting subgroups. In contrast, faultlines are weakest when attributes are not aligned and multiple subgroups can form [23, 328]. To give an example, a faultline is strong in a team consisting of two Caucasian, highly educated women and two African–American men with low level of education. In this case, all three demographic dimensions along which team members differ (race, sex, educational level) split the team along the same line. The faultline would be weaker if, for example, the two highly educated team members would be one man and one woman, and of the two African Americans, one would have a low education and one a high education.

The core prediction of the faultline argument (see [23, p. 331]) is that the negative effects of diversity on performance that research often finds are spurious correlations. Lau and Murnighan propose that not diversity but the strength of the demographic faultline increases dissensus between team members and thus puts performance under pressure. Controlling for faultline strength, the theory implies, direct effects of diversity on performance are always positive because diverse teams have a larger pool of human and social capital that results in higher performance. Faultline strength on the other hand affects performance only negatively.

How do Lau and Murnighan derive this prediction? Their reasoning [23, pp. 332–333] is based on two main mechanisms: *First* they propose that team members prefer to interact with those who are similar with respect to a salient demographic attribute. This corresponds to the widely accepted idea that *homophily* [25] is a strong force in social interactions [28]. Which demographic attribute is salient in a certain work situation changes from situation to situation. *Secondly*, if actors choose to interact they are assumed to exert *social influence* [10] upon each other. Based on psychological research on opinion formation in groups [16,46] Lau and Murnighan propose that social influence occurs in the course of social interactions where team members exchange the arguments their opinions are based on. The authors assume furthermore that demographically similar actors tend to hold similar opinions. The interplay of these mechanisms implies opinion splits in teams with strong faultlines. To explain, if team members hold the same value on a salient demographic attribute then homophily implies that they will more often interact and exchange the arguments their opinions are based on. In the process, all demographically similar interaction partners will become more convinced of their respective opinion, because they tend to agree in opinion and they learn new arguments that are in line with their opinion. In other words their opinions become more extreme.<sup>1</sup> Because in teams with a strong faultline the same team members interact again and again the opinions of the demographic subgroups become more and more distinct which leads to conflicts between the groups at the expense of lower cohesion of the team as a whole.

While Lau and Murnighan’s reasoning is compelling, it remains unclear what exactly the theoretical assumptions are that are needed to derive the prediction that the strength of faultlines has negative effects on cohesion and performance. In other words, the transparency of Lau and Murnighan’s theory would benefit from a formal deduction of their central claims. Moreover, such a formal deduction would facilitate the generation of new hypotheses about conditions under which faultline effects can be expected to occur and possible

<sup>1</sup> Lau and Murnighan need for this reasoning the assumption that demographically similar agents have similar opinions already prior to interaction. However, the authors make this assumption not explicit. To explain emergent subgroup splits, Lau and Murnighan refer to insights from psychological studies [16,46] which show that if interacting actors hold similar opinions, then their opinions become more extreme because they learn from each other new arguments to support their opinions. But this reasoning only implies a subgroup split if the interacting agents already hold similar opinions when they meet. If their prior opinions were dissimilar, then arguments for conflicting opinions would be exchanged through interaction, which in turn would decrease actors’ confidence in their opinions and thus weaken rather than strengthen subgroup splits. We propose in this paper an explanation of faultline effects that does not require the assumption of previous opinion similarity between demographically similar agents.

suggestions for managers for how to prevent negative faultline effects in organizational settings. Accordingly, the objective of the present paper is to fully explicate and formalize the dynamics that may underlie the effects of demographic faultlines on cohesion in demographically diverse groups. For this, we use a formal computational model and show that it explains the effects of faultline strength on cohesion that Lau and Murnighan proposed.

In Section 2, we describe the underlying formal model. Section 3 contains a description of the computational experiments and results. In Section 4, we discuss results, offer conclusions and point to possible future research that could apply our model to generate insights of practical relevance for organizations.

## 2. The model

In this section we develop the formal model. The model consists of three main elements. The first element is the formalization of the dynamics and elementary mechanisms of *social interactions and influence* between team members. The second element consists of our operationalization of the concept of *demographic faultlines*. Thirdly, we devise *aggregate outcome measures* that capture the dependent variables we are interested in.

### 2.1. The social interaction and influence dynamics

We assume that consensus on work related opinions is a major factor for the performance of teams. Accordingly, the main endogenous variable of our model is the *distribution of work related opinions* in the team. Work related opinions may be attitudes about what the team's task is and how to fulfil the task. The distribution of these attitudes can be characterized by the degree to which it resembles one of two theoretical extremes: consensus and polarization. If all team members hold exactly the same opinions then there is perfect *consensus*. If the distribution is bipolar and all team members hold opinions that are maximally different from those on the opposite pole, then this outcome is perfect *polarization*. Obviously, polarization is a major obstacle to good team performance, while consensus at least on fundamental issues seems to be a necessary precondition for effective teamwork.

We focus on the distribution of work related opinions as a main outcome variable for two reasons. First, the team is able to complete its task only if the team members agree to a certain extent on what the task of the team is and how to cope with it [30,32, pp. 175–176]. Second, the theoretical assumptions of homophily and social influence identify a clear causal link between team cohesion, consensus on work related opinions and the strength of demographic faultlines in a team. Broadly, the stronger are faultlines in the team, the less likely it is that team members in different subgroups influence each other sufficiently to generate a consensus on work related opinions on the level of the team as a whole, and the more likely it is that the influence processes result in polarization rather than consensus. At the same time, the combined assumptions of homophily and influence link the degree of consensus closely to the level of cohesion in the team. We assume that only when team members agree on important issues will they have good social relations with each other which, in turn, generates social cohesion.

With this approach, we deliberately exclude from our analysis variables which also may affect performance but which are not or at least much less directly causally related to faultline strength [23], like the size of the team's pool of human and social capital.

What are the model's main ingredients? Our model starts from the assumption that the effects of faultlines on opinion polarization (and poor team performance) are generated by the interplay of the four fundamental social mechanisms *homophily*, *social influence*, *heterophobia* and *rejection*. According to homophily, the more similar two actors are with respect to salient opinions or demographic characteristics, the more they like each other and the more they interact [4,5,15,20,25,28,39]. According to social influence, if two persons interact they adapt their opinions [1,4,21]. Both homophily and social influence are widely documented and robust mechanisms of social interaction and are also used by Lau and Murnighan. In combination, these two mechanisms have a paradoxical implication. The interplay of homophily and social influence implies an inexorable tendency towards consensus. The key reason is a *feed back loop*. Minimal initial similarity leads actors to interact which leads to an increase of similarity which then increases the probability of further interaction. In theory, this process continues until all agents mutually interact and are absolutely similar. However, Axelrod's

(1997) computational studies showed how based on these two mechanisms local convergence can lead to global differentiation. The key assumption that generated this result was *cultural speciation*. Axelrod assumed social interaction between neighbors to be entirely cut off when actors disagree beyond a certain critical level. As a consequence, homophily and social influence stabilize differences between subgroups [2,11,12,14,48].

We believe there are at least two problems with Axelrod's explanation of diversity when it comes to the analysis of opinion polarization in work groups. First, in work groups, there is little room to entirely avoid social interaction with dissimilar others. Both physical exposure and task interdependencies force group members to communicate and take notice of each others' opinions. But even if group members could entirely ignore influence from dissimilar others, Axelrod's model can at best explain why differences between subgroups persist over time. However, it can not readily explain why empirical studies found in groups with strong faultlines a tendency towards extreme and over time increasing opinion differences between a small number of opposed and demographically dissimilar factions in the team (cf. [9]).

To address this pattern, we followed previous research and complemented the mechanisms of homophily and social influence with their negative counterparts of *heterophobia* and *rejection*<sup>2</sup> [17,22,26,34–37,42]. Heterophobia assumes that if the dissimilarity of two actors exceeds a certain threshold then the actors do not like each other [6,8,33,40,41,43]. Rejection states that actors have a tendency to change their attributes in a way to become more dissimilar to interaction partners they do not like [1,22,42,45].

Finally, our model distinguishes between two types of attributes on which agents can differ and which define the level of similarity between agents. Demographic attributes on the one hand are fixed and can not be changed by the dynamics of social influence and rejection. On the other hand, opinions are flexible and are subject to social influence and rejection. Previous computational studies based on similar sets of assumptions have already demonstrated how demographic differences can lead to the emergence of cultural niches in demographic space such that demographically dissimilar actors also hold dissimilar or even radically opposing opinions [26,27]. However, these studies did not address the effects of faultline strength in the demographic distribution.

Technically, each of the  $N$  team members is represented as an agent  $i$  characterized by  $D$  fixed ( $a_{id}^{\text{fix}}$ ) and  $K$  flexible attributes ( $a_{ik}^{\text{flex}}$ ), where  $d$  and  $k$  refer to the  $d$ 'th and  $k$ 'th fixed and flexible attribute, respectively. The fixed attributes correspond to the demographic characteristics; the flexible ones represent the agent's work related opinions. For simplicity, we assume that demographic attributes and opinions are equally salient. Moreover, we focus on clearly distinguishable demographic attributes, expressed by the assumption that demographic attributes are dichotomous and can take either the value  $-1$  or  $+1$  ( $a_{id}^{\text{fix}} \in \{-1; 1\}$ ). Opinions of the team members can instead vary continuously between  $-1$  and  $+1$  ( $-1 \leq a_{ik}^{\text{flex}} \leq +1$ ).

A key assumption of the model is that the direction and strength of influence that an agent  $i$  imposes on an agent  $j$  does not depend directly on the opinion of  $j$ , but it is moderated by the sign and the strength of the interpersonal relation between  $i$  and  $j$ . To model the *interpersonal relations* between the team members we assume a directed graph where  $w_{ij}$  represents the weight of the corresponding relationship ( $-1 \leq w_{ij} \leq +1$ ). If team member  $i$  has contact to team member  $j$  then the weight  $w_{ij}$  takes a value between  $-1$  and  $1$ . A positive weight reflects that  $i$  evaluates  $j$  positively, whereas a negative one represents a hostile relationship. If there is no contact between  $i$  and  $j$ , or  $i$  is indifferent between liking and disliking  $j$ , then the weight is  $0$ .

Both the  $K$  flexible attributes and the weights of the relationships are endogenous and change in discreet time steps. In every time step, one team member is selected randomly with equal probability to update either his flexible attributes or weights. Then, either all weights of  $i$  are updated simultaneously, or all flexible attributes are updated simultaneously, where each option is selected with probability  $0.5$ .

Time is modeled in discreet steps. The duration of a simulation run is expressed in number of iterations. One iteration corresponds to  $N$  simulation steps to assure that on average each agent updates either his weights or his attributes once within an iteration.

Similar to previous models of social influence with continuous opinions [1,14], we assume that the change of team member  $i$ 's flexible attribute  $k$  is an aggregated result of the influences imposed by all other agents who

<sup>2</sup> In Lau and Murnighan's reasoning, increasing opinion differences become possible because the authors assume an initial correlation between demographic attributes and opinions (see also Footnote 1). We discuss in Section 4 why we avoided this assumption.

exert influence upon  $i$  (for an application to organizational enculturation see e.g. [13]). Technically, the new value of the attribute,  $a_{ik,t+1}^{\text{flex}}$  is obtained by adding to the old value a weighted sum of the pressures of all influential others. To model a somewhat gradual change of opinions, we also divide this weighted sum by 2. The pressure imposed by a single alter  $j$  “pulls”  $i$  towards  $j$ ’s opinion if the weight  $w_{ij}$  is positive, and “pushes”  $i$  away from  $j$ ’s opinion if the weight is negative. The magnitude of this pressure is proportional to the distance in opinions between  $i$  and  $j$ ,  $a_{ik}^{\text{flex}} - a_{jk}^{\text{flex}}$ . With only positive weights summing to one, this assumption would imply that the net pressure imposed on  $i$  moves the agent towards the weighted average of the opinions of all interactions partners. Eq. (1) formalizes these assumptions

$$a_{ik,t+1}^{\text{flex}} = a_{ik,t}^{\text{flex}} + \frac{1}{2(N-1)} \sum_{i \neq j} w_{ij} (a_{jk,t}^{\text{flex}} - a_{ik,t}^{\text{flex}}) \quad (1)$$

To be precise, Eq. (1) only shows the principle model of influence. In the actual implementation, we apply a slight modification of the influence equation both to make sure that opinions do not go out of bounds and to smoothen the change of opinions when agents move towards the extreme ends of the opinion scale. Eqs. (1a) and (1b) fully specify opinion change

$$\Delta a_{ik,t}^{\text{flex}} = \frac{1}{2(N-1)} \sum_{j \neq i} w_{ij} (a_{jk,t}^{\text{flex}} - a_{ik,t}^{\text{flex}}) \quad (1a)$$

$$a_{ik,t+1}^{\text{flex}} = \begin{cases} a_{ik,t}^{\text{flex}} + \Delta a_{ik,t}^{\text{flex}} (1 - a_{ik,t}^{\text{flex}}), & \text{if } \Delta a_{ik,t}^{\text{flex}} > 0 \\ a_{ik,t}^{\text{flex}} + \Delta a_{ik,t}^{\text{flex}} (1 + a_{ik,t}^{\text{flex}}), & \text{if } \Delta a_{ik,t}^{\text{flex}} \leq 0 \end{cases} \quad (1b)$$

Eq. (1a) specifies the resulting change of the focal agent’s opinion, given as a weighted sum of the influences imposed by other group members, where the magnitude and direction of the influence depend on the opinion distance  $a_{jk,t}^{\text{flex}} - a_{ik,t}^{\text{flex}}$  weighted by the strength and valence of the relationship,  $w_{ij}$ . Eq. (1b) expresses that the resulting opinion is the sum of the previous opinion and the opinion change times a dampening term that reduces the magnitude of the change to the extent that the opinion moves towards a boundary of the opinion interval. The second key element of our model is the update of weights. Following previous work [26] we assume that the weight that agent  $i$  has towards an agent  $j$ , changes depending on the similarity between  $i$  and  $j$  in terms of both their demographic attributes and their opinions. More precisely, we assume that after updating, the weight adopts a level that is proportional to the current level of similarity. The new weight is negative if the average distance between  $i$  and  $j$  across all dimensions of demographic and opinion space exceeds one, i.e. half of the maximum average distance. If this average distance is exactly one, the weight is zero and otherwise it obtains a positive value. Technically,

$$w_{ij,t+1} = 1 - \frac{\sum_{d=1}^D |a_{id,t}^{\text{fix}} - a_{jd,t}^{\text{fix}}| + \sum_{k=1}^K |a_{ik,t}^{\text{flex}} - a_{jk,t}^{\text{flex}}|}{D + K} \quad (2)$$

## 2.2. Faultline strength

To disentangle the effects of the strength of demographic faultlines from effects of demographic diversity, we devised a method that allows varying faultline strength and keep diversity constant at the same time. More precisely, we generated different distributions of the fixed attributes in such a way that all fixed attributes were equally frequent (all distributions generate equally diverse groups) but the correlation between the attributes differed between distributions (the strength of the faultline differs).

Table 1 shows our construction method for the prototypical case of a group with 20 members ( $N = 20$ ) who differ along three demographic dimensions (e.g. male/female, young/old, western/non-western ethnical background). Column 2 of the table shows that we constructed the first demographic variable ( $A_1$ ) by alternately assigning the values  $-1$  and  $1$  to the first  $N/2$  agents beginning with the value  $1$  for agent 1. We did the same with the second  $N/2$  agents but here we started with the value  $-1$ . The distribution of this variable is the same in all work teams.



Table 1  
Implementation of faultline strength

<i>i</i>	<i>f</i> = 0.9			<i>f</i> = 0.8			<i>f</i> = 0.7			<i>f</i> = 0.6			<i>f</i> = 0.5		
	attr. <i>A</i> <sub>1</sub>	attr. <i>A</i> <sub>2</sub>	attr. <i>A</i> <sub>3</sub>	attr. <i>A</i> <sub>1</sub>	attr. <i>A</i> <sub>2</sub>	attr. <i>A</i> <sub>3</sub>	attr. <i>A</i> <sub>1</sub>	attr. <i>A</i> <sub>2</sub>	attr. <i>A</i> <sub>3</sub>	attr. <i>A</i> <sub>1</sub>	attr. <i>A</i> <sub>2</sub>	attr. <i>A</i> <sub>3</sub>	attr. <i>A</i> <sub>1</sub>	attr. <i>A</i> <sub>2</sub>	attr. <i>A</i> <sub>3</sub>
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1
7	1	1	1	1	1	1	1	1	1	1	1	-1	1	1	-1
8	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	-1	-1	1
9	1	1	1	1	1	-1	1	1	-1	1	1	-1	1	1	-1
10	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1
11	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1
12	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	1
13	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	-1
14	1	1	1	1	1	1	1	1	1	1	-1	1	1	-1	1
15	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	1	-1	-1	1	-1
16	1	1	1	1	1	1	1	-1	1	1	-1	1	1	-1	-1
17	-1	-1	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	1	1
18	1	1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1
19	-1	1	-1	-1	1	1	-1	1	1	-1	1	1	-1	1	1
20	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1
Σ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Correlations ( <i>r</i> )		
	<i>A</i> <sub>2</sub>	<i>A</i> <sub>3</sub>
<i>A</i> <sub>1</sub>	.8	.8
<i>A</i> <sub>2</sub>		.8

	<i>A</i> <sub>2</sub>	<i>A</i> <sub>3</sub>
<i>A</i> <sub>1</sub>	.6	.6
<i>A</i> <sub>2</sub>		.6

	<i>A</i> <sub>2</sub>	<i>A</i> <sub>3</sub>
<i>A</i> <sub>1</sub>	.4	.4
<i>A</i> <sub>2</sub>		.4

	<i>A</i> <sub>2</sub>	<i>A</i> <sub>3</sub>
<i>A</i> <sub>1</sub>	.2	.2
<i>A</i> <sub>2</sub>		.2

	<i>A</i> <sub>2</sub>	<i>A</i> <sub>3</sub>
<i>A</i> <sub>1</sub>	.0	.0
<i>A</i> <sub>2</sub>		.0

We expressed the faultline strength by a parameter  $f$  that varies between 0.5 and 1, where  $f = 0.5$  corresponds to a situation where the demographic attributes are completely uncorrelated and  $f = 1$  imposes a perfect correlation between all demographic attributes. The first step in the construction is to impose the correlation between attribute  $A_1$  and  $A_2$  that corresponds to the given parameter value of  $f$ . To arrive at the values for attribute  $A_2$ , we assigned to the first  $(100 \cdot f)$  % of the cases the same value as for attribute  $A_1$ . This means for example for  $f = 0.9$ , that the first 90% of the agents (the first 18 agents if  $N = 20$ ) hold the same value at attribute  $A_1$  and  $A_2$  (see the grey cells in column 3 of Table 1). To the rest of the agents we assigned on attribute  $A_2$  the opposite value of what we assigned for attribute  $A_1$ .

To determine the values for attribute  $A_3$  we used the same method with a small change. We first assigned to the first  $(50 \cdot f)$  % of the cases the same value as for attribute  $A_1$ . Then we continued with the  $(N/2+1)$ th case and again assigned to the following  $(50 \cdot f)$  % of the cases the same value as for attribute  $A_1$ . Again the rest of the cases got the opposite value than for attribute  $A_1$ . Thus for  $f = 0.9$  and  $N = 20$  the agents 1–9 and 11–19 hold the same value at attribute  $A_1$  and  $A_3$  (see column 4 of Table 1). This procedure makes sure, that the agents also hold at the attributes  $A_2$  and  $A_3$  in exactly  $(100 \cdot f)$  % of all cases the same value.

Table 1 also reports the correlations between the three attributes. Note that for a given distribution all pair-wise correlations between two of the three attributes are equal. The relationship between  $f$  and the correlation is:  $r = -1 + 2f$ . If  $f$  takes the value 1 then the three attributes are perfectly correlated ( $r = 1$ ) and the faultline

strength is maximal. If  $f$  takes the value 0.5 then there is no relationship between the attributes ( $r = 0$ ). Thus the faultline has a minimal strength. At all levels of  $f$ , all variables are equally distributed in all teams. We do not consider in the remainder values of  $f$  below 0.5 (corresponding to negative correlations), because for the strength of a demographic faultline it is only relevant that fixed attributes are correlated, not which direction the correlation takes.

The key advantage of our method is that it separates variation in faultline strength from variation in diversity. A more intuitive alternative approach could have been to assign attributes randomly with a given probability and a given correlation. However, for the relatively small groups we are interested in, that method would have produced considerable random variation in faultline strength between single realizations of distributions imposed by the same level of  $f$ . Our deterministic approach excludes this source of random noise and thus allows us to focus in our computational experiments exclusively on effects of variation in  $f$ .

### 2.3. Aggregate outcome measures

The main claims of the theory of faultlines address two relationships: First the relationship between faultline strength and the level of consensus in the team, and second the relationship between faultline strength and the degree to which divisions in opinions are associated with demographical divisions in the team. To assess whether our model can reproduce these relationships, we devise four different outcome measures. The measure *opinion diversity* captures the number of different positions in the flexible attributes (opinions) represented in the team. *Opinion variance* indicates the average distance between the opinions of different agents. *Polarization* measures the degree to which the team falls apart into a small number of opposed factions who hold maximally different opinions on all dimensions of the opinion space. Finally, we measure the degree to which differences in fixed (demographic) and flexible (opinion) attributes of agents are associated with each other with the outcome measure attribute-opinion covariance,  $cov(fix; flex)$ .

*Opinion diversity* is based on a count of the number of different opinion vectors present in the group as a whole, where only flexible attributes are taken into account. For normalization, we divide this number by the group size  $N$ . Moreover, to express that perfect unity corresponds to no diversity whatsoever, we set *opinion diversity* = 0 if there is perfect consensus.<sup>3</sup> Hence,  $0 \leq \text{opinion diversity} \leq 1$ . We consider the opinion vectors of two agents as different from each other if their flexible attributes  $a_{ik}^{flex}$  differ in at least one dimension  $k$  by a magnitude of 0.001 or more. Clearly, both a group with high consensus and a group with perfect polarization will exhibit low *opinion diversity*. Perfect consensus implies that all agents share the same vector of opinions (*opinion diversity* = zero), whereas perfect polarization implies that there are exactly two maximally different factions in the opinion space (*opinion diversity* =  $2/N$ ).

*Opinion variance* captures a different aspect of consensus than *opinion diversity*. *Opinion diversity* can be high if each agent holds an opinion vector that differs only slightly from that of every other agent, and it can be low if the number of different vectors is small, but the distance between opinions is large, as in the case of perfect polarization. To also represent the magnitude of opinion distances in an outcome measure, we compute *opinion variance* as the average standard deviation of opinions across all  $K$  dimensions of the opinion space. In the case of perfect consensus, we obtain *opinion variance* = 0, and in the case of perfect polarization with two equally large maximally opposed subgroups we measure *opinion variance* = 1, the highest value we ever obtained.<sup>4</sup> However, a high level of *opinion variance* does not necessarily indicate that the group polarizes in the opinion space. High *opinion variance* may occur if agents strongly differ from each other in all dimensions of the opinion space, but these differences are not correlated across dimensions. In that case, the group is not polarized because it is impossible to separate the group into a small number of subsets such that agents

<sup>3</sup> This implies that diversity values do not vary in a continuous interval, but there is a gap between 0 and  $2/N$ . However, this discontinuity is hardly discernible in the results presented below and it does not affect the qualitative interpretation of the measure. We are grateful to J. Richard Harrison for proposing as a better solution  $\text{diversity} = (\text{number of vectors} - 1)/(N - 1)$ , which we will use in future studies.

<sup>4</sup> In this case, the average opinion in all dimensions is zero. Moreover, in all dimensions half of the group adopts an extreme opinion at +1 and the other half of the group does so at -1. Hence, on average the distance from the mean amounts to +1 in all dimensions, yielding the result of *variance* = +1.



maximally agree with each other on all dimensions simultaneously within their subset and disagree on all dimensions with those outside of their subset.

Our measure of *polarization* captures the degree to which the group can be separated into a small set of factions who are mutually antagonistic in the opinion space and have maximal internal agreement. To compute *polarization*, we use the variance of pairwise agreement across all pairs of agents in the population, where agreement is ranging between  $-1$  (total disagreement) and  $+1$  (full agreement), measured as one minus the average distance of opinions (averaged across all  $K$  subdimensions). This measure obviously adopts its lowest level of zero for the case of perfect consensus. The maximum level of polarization (*polarization* = 1) is obtained when the population is equally divided between the opposite ends of the opinion scale at  $-1$  and  $+1$  and all opinion dimensions are perfectly correlated.<sup>5</sup> With uniformly distributed opinions, the polarization measure yields approximately 0.22 for  $K = 1$ .

*Opinion diversity*, *opinion variance* and *polarization* allow us to characterize the degree to which there is polarization or consensus in the opinion distribution of a team. But these measures give no insight into the relationship between demographic differences and differences in opinions. To test this relationship, we compute the attribute-opinion covariance,  $cov(fix; flex)$ . Technically, this index is computed as the covariance between the vector of pairwise demographic dissimilarities and the pairwise opinion dissimilarities, where we computed for every pair of actors  $i$  and  $j$  the dissimilarity measures  $\Delta_{ij}^{fix}$  and  $\Delta_{ij}^{flex}$ , as given by Eqs. (4a) and (4b).

$$\Delta_{i,j}^{fix} = \frac{1}{D} \sum_{d=1}^D |a_{id}^{fix} - a_{jd}^{fix}| \quad (4a)$$

$$\Delta_{i,j}^{flex} = \frac{1}{K} \sum_{k=1}^K |a_{ik}^{flex} - a_{jk}^{flex}| \quad (4b)$$

Thus  $cov(fix; flex)$  is calculated as given by Eq. (5).

$$cov(fix, flex) = \frac{\sum_{j \neq i} \left( (\Delta_{ij}^{fix} - \overline{\Delta^{fix}}) (\Delta_{ij}^{flex} - \overline{\Delta^{flex}}) \right)}{N(N-1)} \quad (5)$$

### 3. Results of the computational experiments

The objective of the computer experiments is to test whether the dynamics of our model are consistent with Lau and Murnighan's [23] informal reasoning. More precisely, we devise a fixed work team scenario and conduct *ceteris paribus* replications of the group dynamics that our model generates for different levels of faultline strength under the given scenario. The stylized regularity our model should produce in this set of experiments is a clear-cut negative relationship between the average level of consensus in the opinion distribution and the strength of demographic faultlines,  $f$ . More in particular, we wish to test whether the model generates both less often consensus and more often polarization as  $f$  increases. A second regularity that follows from the theory of faultline effects and that we want to test is an increasing association of opinion divisions with demographic divisions as faultlines become stronger. In other words, the stronger the demographic faultlines, the clearer we expect subgroup splits in the opinion distribution to reflect the distribution of demographic attributes.

In the experiments we use the following parameter settings. With regard to group size, we assume  $N = 20$ , a size that is not too big to be unrealistic for a work team, but also large enough to allow for a sufficiently fine-grained variation in the strength of demographic faultlines (cf. Table 1). Furthermore, we assume that there are three salient demographic (fixed) attributes ( $D = 3$ ). As Table 1 shows, the combination of 20 agents and 3 fixed attributes allows sufficient variation in the correlations between the fixed attributes of team members. Values for the demographic attributes are assigned to agents as shown in Table 1, imposed by the data set

<sup>5</sup> To see this: In 50% of all dyads the agreement is 1 (indicating maximal agreement), in 50% it is  $-1$  (indicating maximal disagreement). The average level of agreement is zero and the average distance between the agreement in a particular dyad and the average level of agreement, i.e. the variance, yields *polarization* = 1.

we generated for the corresponding level of faultline strength  $f$ . For the number of flexible attributes (opinions), we choose  $K = 4$ . This is the smallest number that makes for  $D = 3$  polarization under strong faultlines not trivial, because with  $K = 4$  and  $D = 3$  (or, more generally,  $K > D$ ) it is still possible that two agents who maximally differ in all three demographic dimensions can have a positive relationship if they have sufficiently similar opinions. At the same time, this setting makes it hard to avoid polarization in a group with maximal faultline strength. Furthermore we assumed that initially (at the outset of  $t = 1$ ) all opinions of all agents are randomly drawn from a uniform distribution with full coverage of the entire opinion interval and with statistically independent dimensions of the opinion space. As a consequence, initial opinions are also statistically independent from demographic attributes, which implies that the expected initial attribute–opinion covariance is zero. After initial opinions have been assigned, initial weights are computed on basis of overall similarity (see Eq. (2)).

To illustrate how variation in faultline strength affects the model dynamics, we show first two typical simulation runs obtained for a setting with low faultline strength ( $r = 0.2$ ) and high faultline strength ( $r = 0.8$ ), respectively. Fig. 1 charts for both settings the dynamics of the four outcome measures for the first 120 iterations.

Fig. 1 shows dramatically different outcomes for the two different levels of faultline strength. In the weak faultline case, the simulated group quickly moves towards perfect consensus, as indicated by the rapid decline of *opinion diversity* and *opinion variance*, as well as *polarization*, from the levels given by the initial random distribution down to the theoretical minimum level of zero for all three outcome measures. The graph also shows that from the outset there is no (actually even a slightly negative) association between differences in opinions and demographic differences (see  $cov(fix; flex)$ ). In the strong faultline case, it takes about 60 iterations until the group has moved from the random initial opinion distribution towards perfect polarization into two maximally opposed factions. Moreover, opinion divisions and demographic divisions align almost perfectly in this case, as indicated by a level of  $cov(fix; flex) = 0.8$  obtained after about 60 iterations.

The explanation for the differences shown by Fig. 1 can be readily derived from our model assumptions. In the weak faultline scenario, demographic attributes are almost perfectly uncorrelated with each other. Hence, there are only very few pairs of agents who maximally differ on all three demographic dimensions. This makes it unlikely that negative ties ( $w_{ij} < 0$ ) arise in the initial configuration. In addition, if some negative ties arise,

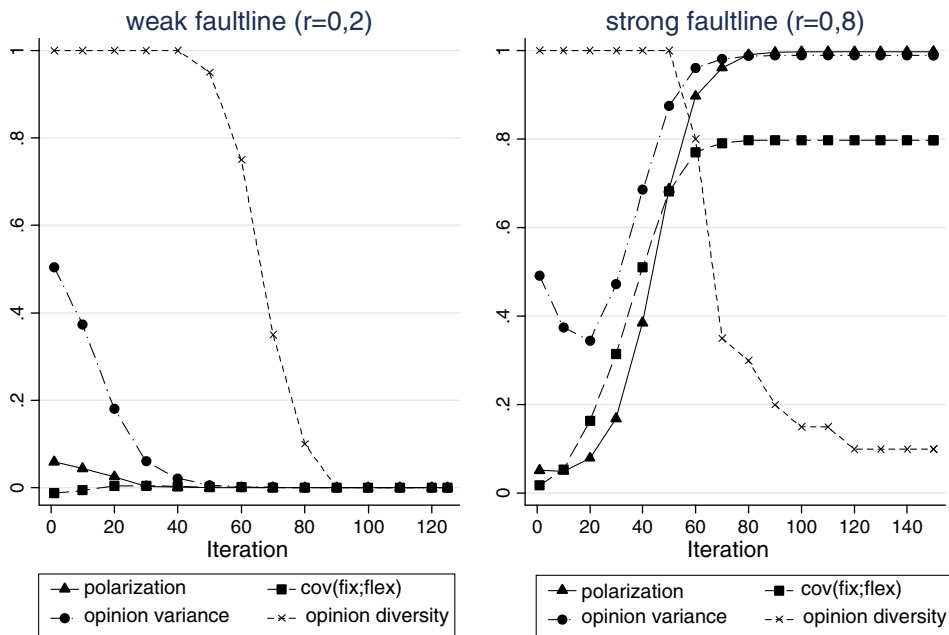


Fig. 1. Change in outcome measure for typical simulation runs with weak faultline (left) and strong faultline (right).  $N = 20$ ,  $D = 3$ ,  $K = 4$ .

then they will most likely be between agents who are, in turn, embedded into a large number of positive ties with the same colleagues. As a consequence, positive social influence prevails and rejection hardly ever occurs in the social interactions between agents. If some agents are “pushed” to reject some enemies’ opinions, then they are at the same time “pulled in” by many more friends so that the net change of their opinion is more likely towards the group average than towards the extreme ends of the opinion scale. A similar reasoning explains why the outcome for the strong faultline case is so different. In the strong faultline case, demographic differences are maximal within a large fraction of the dyads in the team. In these dyads only relatively small opinion differences in the initial configuration suffice to generate a negative relationship between the interactants. Moreover, these negative relationships tend to segregate the two major subgroups in demographic space so that most agents have the same enemies than their friends have. This entails a quick self reinforcing dynamic towards opinion polarization. Most agents move towards whatever is the current average opinion profile in their (demographic) in-group and they distance themselves from whatever is the current average opinion profile in the (demographic) out-group. The result is a coordinated movement of all agents that soon leads to convergence of their opinions on two opposite poles that align with the demographic faultline in the group.

To provide a more detailed representation of the micro level dynamics, we modified the initial conditions such that simultaneous changes of all opinions of all agents can be graphically illustrated. For clarity of the illustration, we assume only one fixed attribute ( $D = 1$ ) along which the group splits into two equally large subgroups with values  $-1$  and  $+1$  respectively. This corresponds to a situation with more than one demographic attributes and a maximally strong faultline. To retain the condition that an emergent opinion polarization is not trivial ( $K > D$ ), we use the smallest number of opinions larger than one that allows an easily interpretable graphic representation,  $K = 2$ . Fig. 2 shows the change of the four outcome measures for a typical run under this parameter constellation. As Fig. 2 shows, the dynamics are very similar to those we find with more demographic attributes and a strong faultline (right part of Fig. 1). *Opinion diversity* decreases and finally takes the value 0.1, indicating that at the end of the simulation run there are only two different opinion vectors left in the team. All other outcome measures approach relatively fast the maximal value 1. This shows that members of the two subgroups hold maximally different values on the two opinion dimensions and that the two subgroups also differ in the demographic attribute. The maximal level of *cov(fix, flex)* in particular indicates that the demographic attribute and the opinions are perfectly correlated at the end of the process.

Fig. 3 gives insight into the micro dynamics of the simulation run reported in Fig. 2. It shows scatter plots of the distribution of opinions in the two dimensional opinion space. The circles and crosses indicate the

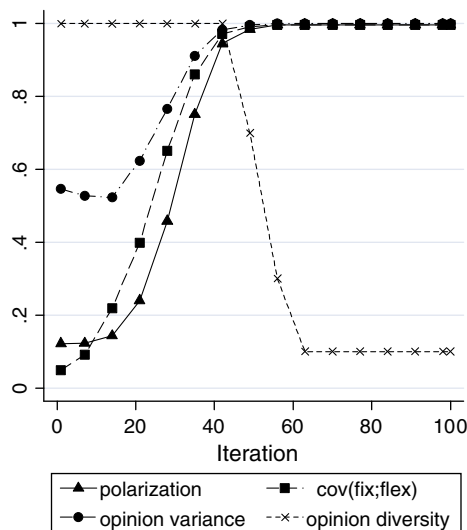


Fig. 2. Change in outcome measure for typical simulation run.  $N = 20$ ,  $D = 1$ ,  $K = 2$ .

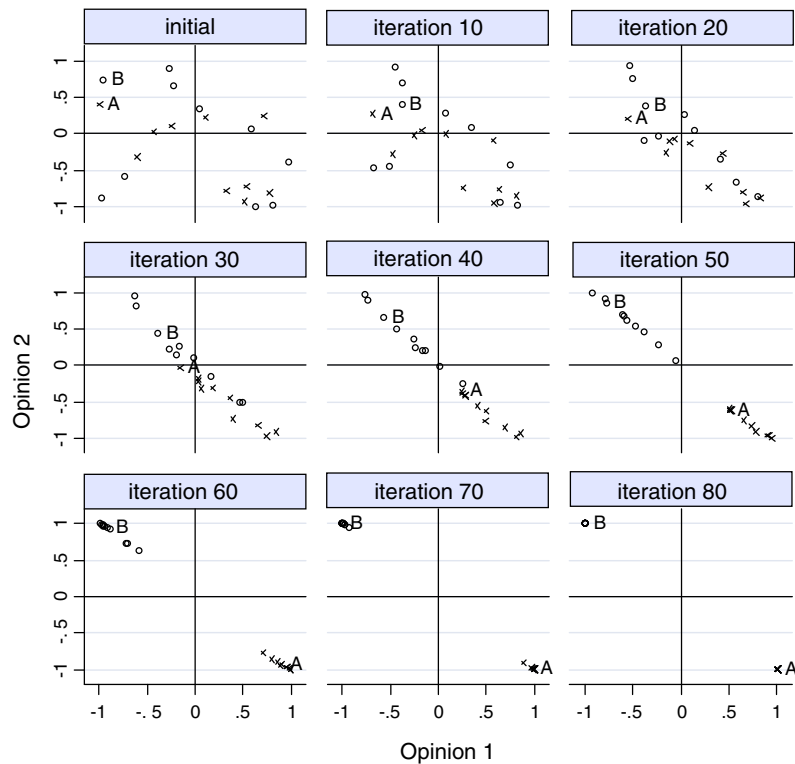


Fig. 3. Development of the opinion distribution for typical simulation run.  $N = 20$ ,  $D = 1$ ,  $K = 2$ .

demographic attribute of the respective team member – thus there are 10 circles and 10 crosses. We labelled two agents (A and B) with different demographic attribute to highlight their opinion changes. The top left scatter plot shows the initial randomly assigned opinion distribution. In the early phase of the team process some of the agents approach more moderate opinions (see also the slight decline of opinion variance in the first iterations reported in Fig. 2). For example agent B, who initially held quite extreme opinions on both dimensions, holds less extreme opinions in iterations 10 and 20 than in the initial condition. The reason is that in the initial condition, agent B has some overlap with many of the other team members, most of whom hold opinion positions relatively close to the center. As a consequence, agent B has positive relations to most of these other group members. Agent B thus experiences initially positive influence by others and becomes more similar to these others. Because B initially holds relatively extreme opinions, he also becomes more moderate in the process.

However there are also some agents who either keep their extreme positions or develop even more extreme opinions (see the two circles next to B and the crosses with positive values at opinion one and negative values at opinion two in the initial opinion distribution). Those agents are negatively influenced by relatively many others who are dissimilar with respect to both the demographic attribute and the opinions (heterophobia). As a consequence, they approach opinions that differ from the opinions of those actors (rejection). That is, their opinions become more extreme. Subsequently, emergent “extremists” pull those team members who are relatively similar to them also to the extremes.

From iteration 20 on the developing subgroup split becomes obvious. While the circles tend to hold a negative value at opinion one and a positive at opinion two, the crosses tend to hold positive values at opinion one and negative values at opinion two. As a consequence the team members have positive relationships to demographically similar actors and dislike dissimilar team members. This leads to further increasing opinion differences. From iteration 50 on all circles hold negative values at opinion one and negative values at opinions two. The crosses hold opinions with opposing signs. At about iteration 80 equilibrium is reached. The team is now perfectly polarized.

Fig. 3 also shows that agents after some time align on the diagonal of the opinion space. The reason is that agents update their opinions separately per dimension by taking the *mean* of opinion differences of all team members on this dimension (see Eq. (1)). At the same time, the more similar are the opinions of two agents across all dimensions, the more these agents like each other (see Eq. (2)) and the stronger is their influence on each other (see  $w_{ij}$  Eq. (1)). Hence on both dimensions of the opinion space agents tend to adopt opinions that align the distance of their opinion to the group profile on the focal dimension with their overall relational weights vis a vis the other group members. For example, “circle agents” who are relatively close to the extreme opinion vector  $(-1, +1)$  are highly likely to have negative relations to most “cross agents” and will thus in both opinion dimensions move away from the average opinion of all “cross agents” and move towards the average opinion of their fellow circle agents. Less extreme circle agents will distance themselves only from a smaller fraction of the population and thus are more likely to hold – at least temporarily – opinions closer to the centre. As this happens on both opinion dimensions, agents will in both dimension adopt about the same distance from the average group profile, which results in a gradual move towards the diagonal of the opinion space. Which of the two diagonals of the two dimensional opinions space the actors coordinate on depends on how the opinions are distributed initially. In the run shown in Fig. 3 most actors initially hold opinion combinations in the second and fourth quadrant of the opinions space. Accordingly the actors coordinate on the diagonal that is located in these two quadrants. We also found numerous runs where actors coordinated on the other diagonal of the opinion space.

We chose the simple case of  $D = 1$  and  $K = 2$  only for sake of illustration. For statistical reliability, we conducted a large number of replications of the first computational experiment (with three demographic and four opinion dimensions, i.e.  $D = 3$ ,  $K = 4$ ) and varied faultline strength across the entire interval between  $r = 0$  and  $r = 1.0$  in steps of 0.2. Fig. 4 reports the average of the outcome measures we obtained after iteration 1000, over 500 replications per condition. We do not report *opinion diversity* in Fig. 4, because for all levels of faultline strength, final states are almost always either perfectly polarized or exhibit perfect consensus so that the variation of *opinion diversity* across conditions is extremely small. A detailed representation of the opinion distributions for all conditions is given in Fig. 5.

Fig. 4 clearly confirms that our model generates the stylized regularities predicted by Lau and Murnighan’s theory of faultlines. All three outcome measures consistently increase with higher levels of faultline strength.

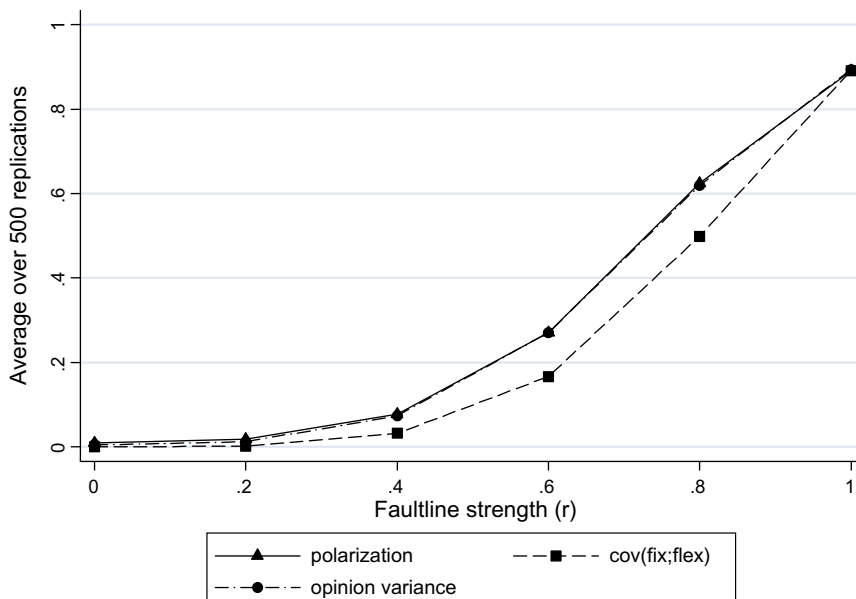


Fig. 4. Effect of faultline strength on outcome measures, averages over 500 replications per conditions, outcomes measured after 1000 iterations per replication  $N = 20$ ,  $D = 3$ ,  $K = 4$ .

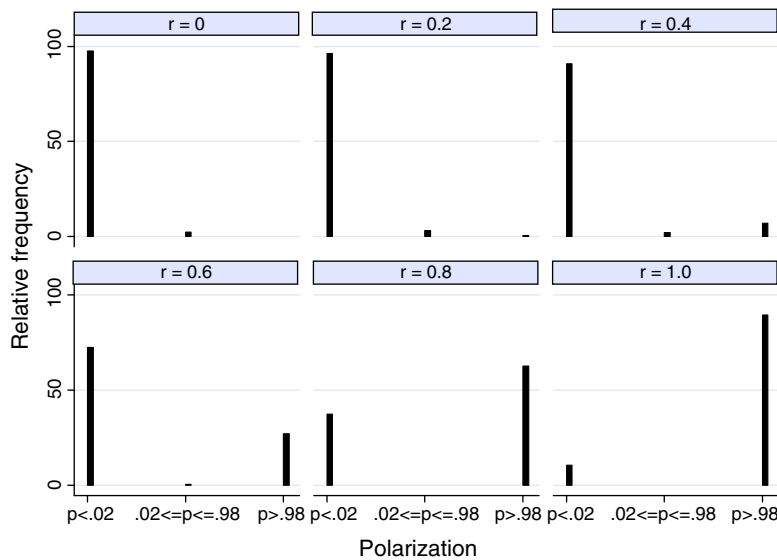


Fig. 5. Distribution of polarization measure over 500 replications per condition in experiment 1, broken down by the six different levels of faultline strength.  $N = 20$ ,  $D = 3$ ,  $K = 4$ .

More specifically, the average outcomes of almost zero for *opinion variance*, *polarization* and *cov(fix; flex)* when demographic dimensions are entirely unrelated ( $r = 0$ ) indicate that virtually all simulated groups have reached almost perfect consensus in this condition. By contrast, with maximal faultline strength ( $r = 1.0$ ) groups almost always polarize maximally, as indicated by an average *polarization* and an average *opinion variance* at the same level. The correspondingly high value of *cov(fix; flex)* in this condition shows that it is the demographic faultline along which the group also splits in the opinion space. The consistent increase of the outcome in between these two extremes shows that – for the given set of conditions ( $N = 20$ ,  $D = 3$ ,  $K = 4$ ) – our model clearly implies that higher faultline strength is associated with less consensus, more polarization and a stronger association between demographic and attitudinal differences, as predicted by Lau and Murnighan's theory.

Fig. 4 shows that faultline strength affects the average outcome measures in the expected way, but it gives no insight into the distribution of qualitatively different end states that occur in the population of groups that were simulated under the same condition. Fig. 5 shows this distribution. The figure charts for six different levels of faultline strength the relative frequencies of the *polarization* index generated across the 500 groups simulated under the corresponding condition.

Fig. 5 reveals that model dynamics almost always converge on one of two extreme outcomes, perfect consensus (*polarization* < 0.02) or perfect polarization (*polarization* > 0.98). This shows that the gradual shift of average outcome measures reported in Fig. 4 for increasing faultline strength is mainly generated by a shift of the distribution between these two extremes. At  $r = 0$ , about 98% of all replications generated perfect consensus, while at  $r = 1.0$  perfect polarization was obtained for roughly 89% of the sample. The most balanced distribution arises at a faultline strength of  $r = 0.8$ , where about 37% of the runs yielded consensus and the remaining 63% produced polarization. The explanation for this result is that perfect consensus and perfect polarization both constitute robust equilibria of the system. Once perfect consensus is reached, no group member is pushed or pulled to change any of her opinions and all weights are at their maximum value of +1. In a similar vein, once perfect polarization is reached, all “friends” with equal opinions reinforce an agent to stay the course, but so do all the enemies with maximally different opinions. Both outcomes are equilibria independently of the distribution of demographic variables that we simulated, because with some appropriate distribution of the four flexible attributes it is always possible to either impose positive ties between all dyads (consensus) or split the group into two subgroups where all in-group ties are positive and all out-group ties are negative. Moreover, these equilibria are not the only possible equilibria. In principle, multiplex equilibria



can also arise in which more than two different subgroups form in the opinion space such that the overall pattern of relationships and opinions is exactly balanced so that “push” and “pull” forces exerted upon agents’ opinion from different groups of friends and enemies exactly neutralize each other (cf. [26]).

While perfect consensus, perfect polarization and multiplex opinion distribution can all be equilibrium outcomes under all simulated conditions, the level of faultline strength strongly affects the likelihood that a particular one of the possible equilibria is reached by the dynamics of the system. To begin with, the complexity of the coordination of moves that is needed to obtain a multiplex equilibrium is very high and thus it is under all conditions quite unlikely that the model generates a multiplex outcome from a random start. This is consistent with the extremely low frequency of multiplex outcomes in the final states of the simulation as reported in Fig. 5 (see the cases with a final polarization value higher than 0.2 and smaller than 0.98) of at most a few percent of all replications conducted. Beyond this, the stronger the faultline, the “easier” it is for a group to coordinate on a sequence of interactions that end up in perfect polarization, because stronger faultlines come with more negative ties and a higher alignment of negative ties with demographic subgroup boundaries. Conversely, the weaker the faultline, the more likely it is that there is a sufficiently large concentration of positive ties to let all agents move towards a consensus in the opinion space.

#### 4. Summary and discussion

We modeled in this paper the effects of demographic faultlines on team performance. Lau and Murnighan’s theory suggests that the stronger a team’s demographic faultline is, the less cohesive the team will be and the less likely will the team therefore be able to find a consensus with regard to work related opinions. As a consequence, teams with a strong demographic faultline tend to perform poorly. However, Lau and Murnighan’s reasoning is not sufficiently explicit to allow a detailed theoretical analysis of the conditions under which these effects may occur. The aim of our paper was to provide an explication of the theory in terms of a formal computational model. Our model is based on four fundamental social mechanisms, homophily, heterophobia, social influence and rejection. We found that the model generates results that are consistent with Lau and Murnighan’s faultline theory. Our computational experiments show that the stronger the demographic faultline in a group the more likely will the group split up into subgroups (*ceteris paribus*). In this event, members of different subgroups hold opposing opinions, differ in their demographic attributes and do not like each other.

This paper focused on Lau and Murnighan’s hypothesis that demographic faultlines affect the chance that work teams reach consensus on work related opinions. However we neglected their proposition that demographic diversity has no direct effects on opinion dynamics. In all experiments reported here we kept diversity maximal (50/50 distributions of all demographic attributes) and only varied the strength of the faultline. It turned out that if there was no faultline or the faultline was very weak then most teams reached consensus although diversity was maximal. This indicates that also Lau and Murnighan’s proposition about the effects of demographic diversity is in line with our model. However future research should also study our model’s predictions if the demographic diversity is varied.

Future research should furthermore focus on the mechanisms that produce opinion polarization. As we argued above it is not possible to explain opinion polarization by using the homophily and social influence mechanisms alone. Further assumptions, like Axelrod’s assumption of cultural specification, have to be included into the model. In Lau and Murnighan’s reasoning opinion polarization becomes possible because an initial correlation between the demographic variables and the opinions is assumed. The mechanism that Lau and Murnighan seem to assume is that due to homophily, demographically similar actors tend to interact. If demographic attributes and opinions correlate then interacting agents tend to agree in opinions. Once they exchange different arguments for the same opinions, their opinions become even more extreme. As a consequence, the difference between the opinions of the demographic subgroups increases and perfect polarization becomes possible. Without the assumption of the initial correlation between demographic attributes and opinions, there is no reason to assume initial agreement of interaction partners. But then the theory does not also imply that the exchange of arguments will render interaction partners more extreme in their opinions. Instead, the prediction would be that actors tend to adopt more moderate opinions. Thus, with Lau and Murnighan’s

original mechanism, the effect of strong demographic faultlines on opinion polarization is hard to explain without the assumption that demographic attributes already split the group to some extent in the first place.

In our model, on the other hand, the two negative mechanisms of heterophobia and rejection are sufficient to generate an effect of faultline strength on opinion polarization. Why did we decide not to use the Lau and Murnighan way of explaining polarization? The main argument is that the assumption of correlating demographic attributes and opinions should not be used to explain the effects of faultline strength on opinion polarization because then one includes the explanandum already in the explanans. What we try to explain is that the strength of the demographic faultline leads to opinion polarization along this faultline. If we already assume in the model that demographic attributes are correlated with the opinions then it is not surprising that the model predicts exactly this as an outcome. Hence, we argue that the assumption of an initial positive correlation between demographic attributes and opinions should be avoided in this context. However, our argument is purely theoretical. We do not claim that the dynamics that Lau and Murnighan describe do not occur in real work teams. This remains an empirical question.

The main advantage of a formal model is that it facilitates the deduction of new hypotheses. We believe that in future research our model can be used to generate new theoretical propositions about the conditions under which faultline effects may occur. One example is the possible proposition that not only faultlines in the distribution of demographic attributes, but also faultlines in the distribution of initial opinions, or a combination of both may affect the chances that a group split emerges. Our model suggests that an initial tendency of the group to fall apart into a small number of opposed subgroup in terms of opinions may set off a self-reinforcing dynamic in which eventually group polarization prevails even when there may not be a strong demographic faultline. Another interesting possible implication is that the effects of faultline strength depend on the timing of contacts, that is on *who* is *when* brought in contact with *whom*.

The key reason why we expect the timing of contacts to be important is the inherent path dependence of the dynamics of social interactions between team members. For example, early contacts between group members who are strongly dissimilar both in terms of their opinions and their demographic characteristics may trigger negative and hostile interactions between the interactants. This, in turn, may lead them to adopt extreme positions on some issues. If these “radicalized” actors interact subsequently with demographically similar “friends”, this may entail “bandwagon dynamics” in which the friends of the early conflict partners are socially influenced to adopt similarly extreme positions. The stronger the demographic faultlines, the more such a dynamic would project the demographic faultline onto an emergent faultline in the opinion space, with the result that communication between team members and thus group cohesion and team performance may severely suffer. Clearly, this downward spiral might be avoided when contacts between team members are arranged in such a way that opposed “extremists” are initially isolated from each other and are instead exposed to interactions with demographically similar in-group members who are more moderate in their opinions. Then, the likely consequence is that initial extremists also become more moderate and initial moderates from different demographic subgroups move towards each other in the opinion space.

In a nutshell this leads to the somewhat counterintuitive prediction that cohesion in teams with a strong faultline can be increased if the team is first separated into two demographically homogeneous subgroups and only later merged. We suggest that future research should focus on the timing of contacts because this may give managers a handle to temper the negative effects of demographic faultlines on team cohesion and performance. We believe that the model we have proposed in this paper is a useful theoretical tool to explore this possibility.

## Acknowledgements

The authors thank Károly Takács, Tobias Stark and the other members of the discussion group on norms and networks at the University of Groningen, as well as J. Richard Harrison for inspiring feedback and comments on previous versions of this chapter. We gratefully acknowledge financial support by the Netherlands Organization for Scientific Research (NWO) under the Innovational Research Incentive Scheme (VIDI Grant #452-04-351).

## References

- [1] R.P. Abelson, Mathematical models of the distribution of attitudes under controversy, in: N. Frederiksen, H. Gulliksen (Eds.), *Contributions to Mathematical Psychology*, Rinehart Winston, New York, 1964, pp. 142–160.
- [2] R. Axelrod, The dissemination of culture – a model with local convergence and global polarization, *Journal of Conflict Resolution* 41 (1997) 203–226.
- [3] C.A. Bowers, J.A. Pharmer, E. Salas, When member homogeneity is needed in work teams. A meta-analysis, *Small Group Research* 31 (2000) 305–327.
- [4] D.J. Brass, J. Galaskiewicz, H.R. Greve, W. Tsai, Taking stock of networks and organizations: a multilevel perspective, *Academy of Management Journal* 47 (2004) 795–817.
- [5] D. Byrne, *The Attraction Paradigm*, Academic Press, New York, London, 1971.
- [6] D. Byrne, G.L. Clore, G. Smeaton, The attraction hypothesis: do similar attitudes affect anything?, *Journal of Personality and Social Psychology* 51 (1986) 1167–1170.
- [7] J.A. Chatman, J.T. Polzer, S.G. Barsade, M.A. Neale, Being different yet feeling similar: the influence of demographic composition and organizational culture on work process and outcome, *Administrative Science Quarterly* 43 (1998) 749–780.
- [8] F.F. Chen, D.T. Kenrick, Repulsion or attraction? Group membership and assumed attitude similarity, *Journal of Personality and Social Psychology* 83 (2002) 111–125.
- [9] P.C. Early, E. Mosakowski, Creating hybrid team cultures: an empirical test of transnational team functioning, *Academy of Management Journal* 43 (2000) 26–49.
- [10] L. Festinger, S. Schachter, K. Back, *Social Pressures in Informal Groups*, Stanford University Press, Stanford, CA, 1950.
- [11] A. Flache, M.W. Macy, Why more contact may increase cultural polarization, paper presented in Session Mathematical Sociology I at 101st Annual Meeting of the American Sociological Association, Montreal, 2006, Reprint published at [www.arXiv.org>physics/0604196](http://www.arXiv.org>physics/0604196).
- [12] A. Flache, M.W. Macy, K. Takács. What sustains stable cultural diversity and what undermines it? Axelrod and Beyond, paper presented in World Congress of Social Simulation 2006, Kyoto, Japan, 2006.
- [13] J.R. Harrison, G. Carroll, The dynamics of cultural influence networks, *Computational and Mathematical Organization Theory* 8 (2002) 5–30.
- [14] R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence models, analysis, and simulation, *Journal of Artificial Societies and Social Simulation* 5 (2002). Available from: <http://jasss.soc.surrey.ac.uk/5/3/2.html>.
- [15] G.C. Homans, *The Human Group*, Harcourt Press, New York, 1951.
- [16] D.J. Isenberg, Group polarization: a critical review and meta-analysis, *Journal of Personality and Social Psychology* 50 (1986) 1141–1151.
- [17] W. Jager, F. Amblard, Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change, *Computational and Mathematical Organization Theory* 10 (2005) 295–303.
- [18] K.A. Jehn, K. Bezrukova, A field study of group diversity, workgroup context, and performance, *Journal of Organizational Behavior* 25 (2004) 703–729.
- [19] K.A. Jehn, G.B. Northcraft, M.A. Neale, Why differences make a difference: a field study of diversity, conflict, and performance in workgroups, *Administrative Science Quarterly* 44 (1999) 741–763.
- [20] D.B. Kandel, Homophily, selection, and socialization in adolescent friendships, *American Journal of Sociology* 84 (1978) 427–436.
- [21] N.L. Kerr, S.R. Tindale, Group performance and decision making, *Annual Review of Psychology* 55 (2004) 623–655.
- [22] J. Kitts, Social influence and the emergence of norms amid ties of amity and enmity, *Simulation Modelling Practice and Theory* 14 (2006) 407–422.
- [23] D.C. Lau, J.K. Murnighan, Demographic diversity and faultlines: the decompositional dynamics of organizational groups, *Academy of Management Review* 23 (1998) 325–340.
- [24] D.C. Lau, J.K. Murnighan, Interactions within groups and subgroups: the effects of demographic faultlines, *Academy of Management Journal* 48 (2005) 645–659.
- [25] P.F. Lazarsfeld, R.K. Merton, Friendship and social process: a substantive and methodological analysis, in: M. Berger, T. Abel, C.H. Page (Eds.), *Freedom and Control in Modern Society*, Van Nostrand, New York, Toronto, London, 1954, pp. 18–66.
- [26] M.W. Macy, J. Kitts, A. Flache, S. Benard, Polarization and dynamic networks. A Hopfield model of emergent structure, in: R. Breiger, K. Carley, P. Pattison (Eds.), *Dynamic Social Network Modelling and Analysis: Workshop Summary and Papers*, The National Academies Press, Washington, DC, 2003, pp. 162–173.
- [27] N.P. Mark, Culture and competition: homophily and distancing explanations for cultural niches, *American Sociological Review* 68 (2003) 319–345.
- [28] M. McPherson, L. Smith-Lovin, J.M. Cook, Birds of a feather: homophily in social networks, *Annual Review of Sociology* 27 (2001) 415–444.
- [29] F.J. Milliken, L.L. Martins, Searching for common threads: understanding the multiple effects of diversity in organizational groups, *Academy of Management Journal* 21 (1996) 402–433.
- [30] E. Molleman, Diversity in demographic characteristics, abilities and personality traits: do faultlines affect team functioning?, *Group Decision and Negotiation* 14 (2005) 173–193.
- [31] L.H. Pelled, Demographic diversity, conflict, and work group outcomes: an intervening process theory, *Organization Science* 7 (1996) 615–631.
- [32] J. Pfeffer, Organizational demography: implications for management, *California Management Review* 28 (1985) 67–81.

- [33] N.W. Pilkington, J.E. Lydon, The relative effect of attitude similarity and attitude dissimilarity on interpersonal attraction: investigating the moderating roles of prejudice and group membership, *Personality and Social Psychology Bulletin* 23 (1997) 107–116.
- [34] K. Rainio, *A Stochastic Model of Social Interaction*, Munksgaard, Turku, 1961.
- [35] K. Rainio, Stochastic process of social interaction, *Scandinavian Journal of Psychology* 2 (1961) 113–128.
- [36] K. Rainio, *A Stochastic Theory of Social Contacts. A Laboratory Study and an Application to Sociometry*, Munksgaard, Turku, 1962.
- [37] K. Rainio, Social interaction as a stochastic learning process, *Archives Europeennes de Sociologie* 6 (1965) 68–88.
- [38] R. Reagans, E.W. Zuckerman, Networks, diversity, and productivity: the social capital of corporate R& D teams, *Organization Science* 12 (2001) 502–517.
- [39] E.M. Rogers, D.K. Bhowmik, Homophily–Heterophily: relational concepts for communication research, *Public Opinion Quarterly* 34 (1970) 523–538.
- [40] M.E. Rosenbaum, Comment on a proposed two-stage theory of relationship formation: first, repulsion; then, attraction, *Journal of Personality and Social Psychology* 51 (1986) 1171–1172.
- [41] M.E. Rosenbaum, The repulsion hypothesis: on the nondevelopment of relationships, *Journal of Personality and Social Psychology* 51 (1986) 1156–1166.
- [42] L. Salzarulo, A continuous opinion dynamics model based on the principle of meta-contrast, *Journal of Artificial Societies and Social Simulation* 9 (2006). Available from: <<http://jasss.soc.surrey.ac.uk/9/1/13.html>>.
- [43] G. Smeaton, D. Byrne, S.K. Murnen, The repulsion hypothesis revisited: similarity irrelevance or dissimilarity bias?, *Journal of Personality and Social Psychology* 56 (1989) 54–59.
- [44] G.L. Stewart, A meta-analytic review of relationships between team design features and team performance, *Journal of Management* 32 (2006) 29–54.
- [45] R. Tsuji, Interpersonal influence and attitude change toward conformity in small groups: a social psychological model, *Journal of Mathematical Sociology* 26 (2002) 17–34.
- [46] A. Vinokur, E. Burnstein, Depolarization of attitudes in groups, *Journal of Personality and Social Psychology* 36 (1978) 872–885.
- [47] S.S. Webber, L.M. Donahue, Impact of highly and less job-related diversity on work group cohesion and performance: a meta-analysis, *Journal of Management* 27 (2001) 141–162.
- [48] G. Weisbuch, G. Deffuant, F. Amblard, Persuasion dynamics, *Physica A* 353 (2005) 555–575.
- [49] K.Y. Williams, C.A. O'Reilly, Demography and diversity in organizations. A review of 40 years of research, *Research in Organizational Behavior* 20 (1998) 77–140.